

Optimal Estimation of Vanishing Points in a Manhattan World

Faraz M. Mirzaei and Stergios I. Roumeliotis*

University of Minnesota
Minneapolis, MN 55414, USA
{faraz|stergios}@cs.umn.edu

Abstract

In this paper, we present an analytical method for computing the globally optimal estimates of orthogonal vanishing points in a “Manhattan world” with a calibrated camera. We formulate this as constrained least-squares problem whose optimality conditions form a multivariate polynomial system. We solve this system analytically to compute all the critical points of the least-squares cost function, and hence the global minimum, i.e., the globally optimal estimate for the orthogonal vanishing points. The same optimal estimator is used in conjunction with RANSAC to generate orthogonal-vanishing-point hypotheses (from triplets of lines) and thus classify lines into parallel and mutually orthogonal groups. The proposed method is validated experimentally on the York Urban Database.

1. Introduction

Estimation of a camera’s orientation in man-made environments is essential in many applications, such as personal navigation and place recognition. It is well known that in the so-called Manhattan world [7], where the 3D lines are aligned with the cardinal axes of the global frame, the noise-free vanishing points are the scaled rows of the rotation matrix representing the camera’s orientation with respect to the global frame [12]. This relationship has been commonly exploited to estimate a calibrated camera’s orientation from vanishing points.

In this work, we consider the problem of estimating orientation and vanishing points of an intrinsically calibrated camera in a Manhattan world, where the line directions are predominantly orthogonal to each other. The state-of-the-art methods for this task [9, 23] are iterative, require accurate initialization, and are not guaranteed to converge to the global optimum. Additionally, existing initialization methods [1, 21] do not enforce orthogonality of the vanishing

points, and thus may not produce sufficiently accurate estimates. To address these limitations, we propose a novel method for analytically determining the optimal orthogonal vanishing points, and an efficient RANSAC-based line classifier to group lines into parallel and mutually orthogonal sets. In summary, the contributions of this work are:

- A globally optimal estimator for vanishing points that (i) is not iterative and does not require any initialization, (ii) is guaranteed to find all globally optimal estimates of the orthogonal vanishing points and the camera’s orientation (in a least-squares sense), (iii) can work with as few as three lines (the minimal problem) or as many as hundreds of lines, and its computational complexity is only linear in the number of lines.
- An efficient RANSAC-based line classifier that uses triplets of lines to generate hypotheses for all three orthogonal vanishing points at once. This RANSAC algorithm works robustly with very few sample measurements, and does not require a dominant line direction.

2. Previous Work

The early works on vanishing-point estimation relied on the Hough transform of the line segments on the Gaussian sphere [4, 14]. These approaches, however, are not reliable in the presence of noise and outliers and may miss-classify lines into incorrect parallel groups [19]. Furthermore, these methods do not typically enforce the orthogonality constraints between the vanishing points leading to suboptimal estimation of both the vanishing points and the camera’s orientation. The exhaustive search method of Rother [17] enforces the orthogonality of the vanishing points. However, its computational complexity is prohibitive for real-time applications.

To address the miss-classification and optimality issues, Expectation-Maximization (EM)-based methods assign a probability to each line segment or image region, indicating the likelihood that it belongs to each of the parallel line groups or an outlier group (expectation step); then, they find the most likely vanishing points from the line assignments (maximization step) [2, 12, 7, 18, 23, 9]. The EM

*This work was supported by the University of Minnesota (DTC), and the National Science Foundation (IIS-0643680, IIS-0811946, IIS-0835637).

approaches suffer from two common drawbacks: (i) they are iterative in nature, and sensitive to initialization, (ii) the maximization step entails optimization of a nonconvex cost function. The EM is typically initialized using results of the Hough transform or heuristic clustering of line-segment intersections, which are not guaranteed to produce a sufficiently accurate initialization. Additionally, the maximization step is usually performed using iterative algorithms, such as gradient descent, which are not guaranteed to converge to the global optimum, and may not find all global optima, if more than one exists.

More recently, RANSAC-based algorithms have been proposed that consider intersections of line segments as hypotheses for vanishing points and prune improbable hypotheses using heuristic criteria [1, 23]. In [10], such RANSAC hypotheses are used to initialize an iterative maximum-likelihood estimator of the vanishing points. In [21], a J-Linkage algorithm is used to generate hypothetical classes of parallel lines, followed by EM to find the vanishing points. These methods, however, do not enforce orthogonality of the vanishing points when generating the hypotheses, and generally require a large number of line segments and sample hypotheses to converge to the correct solution.

In order to address these limitations, in this paper we first introduce a novel algorithm to estimate the (calibrated) camera's orientation and the orthogonal vanishing points, and then propose a RANSAC classifier that uses the proposed estimator to partition line segments into parallel groups (and an outlier group). Specifically, we formulate a least-squares estimator for the camera's orientation and the orthogonal vanishing points whose optimality conditions form a system of polynomial equations in the orientation of the camera. Using tools from algebraic geometry, we solve this system to find the finite set of all the critical points of the least-squares cost function, and choose the one(s) that minimize it. The orthogonal vanishing points are then readily computed as functions of the camera's orientation. The developed algorithm is not iterative, does not require any initialization, and is guaranteed to find the globally optimal estimates of the camera's orientation and the orthogonal vanishing points. Furthermore, the proposed algorithm can work with as few as three lines (the minimal case) or as many as hundreds of lines, with linear computational complexity in the number of lines.

3. Technical Approach

3.1. Notation

Throughout this paper, ${}^A\mathbf{x}$ denotes the expression of a vector \mathbf{x} with respect to frame $\{A\}$ and ${}^A_B\mathbf{C}$ is the rotation matrix rotating vectors from frame $\{B\}$ to frame $\{A\}$. $\bar{\mathbf{x}}$ is a unit vector, $\hat{\mathbf{x}}$ is the estimate of the quantity \mathbf{x} , \mathbf{I}_n

is the $n \times n$ identity matrix, and $\mathbf{0}_{m \times n}$ is the $m \times n$ matrix of zeros. In the representation of a polynomial function $f(\mathbf{x})$ we may drop \mathbf{x} if that does not result in ambiguity. Lines are parametrized with Plücker coordinates consisting of a 3×1 direction vector ${}^A\bar{\mathbf{n}}$ and a 3×1 moment vector ${}^A\mathbf{m} \triangleq {}^A\mathbf{p} \times {}^A\bar{\mathbf{n}}$ where ${}^A\mathbf{p}$ is any point on the line. It follows from the moment definition that ${}^A\mathbf{m}^T {}^A\bar{\mathbf{n}} = 0$. The direction of a moment vector uniquely specifies a plane (called the moment plane) that passes through the line and the origin of the frame of reference.

We assume that the camera is intrinsically calibrated, and thus consider it to follow a spherical projection model, measuring 3D lines as the intersection of the lines' moment planes (in the camera frame) with their Gaussian (unit) sphere. These intersections are great circles that are uniquely determined by the *direction* of the lines' moment planes in the camera's frame. Therefore, we simply assume that the camera measures the direction of the moment planes in its frame, i.e., ${}^C\bar{\mathbf{m}} = \frac{{}^C\mathbf{m}}{\|{}^C\mathbf{m}\|}$.

3.2. Vanishing Points

In a calibrated camera, a vanishing point associated with a line is the direction of the line in the camera's frame of reference, i.e., $\bar{\mathbf{v}} \triangleq {}^C\bar{\mathbf{n}}$. By definition, the vanishing point lies on the moment plane, i.e., $\bar{\mathbf{v}}^T {}^C\bar{\mathbf{m}} = 0$. Note that parallel lines have the same direction, and thus share the same vanishing point which corresponds to the intersection of their moment planes. In a Manhattan world, where 3D lines can be partitioned into three parallel and mutually orthogonal groups, the three vanishing points found in the image are also mutually orthogonal. In other words, if we denote them with $\bar{\mathbf{v}}_i$, $i = 1, 2, 3$, then $\bar{\mathbf{v}}_1^T \bar{\mathbf{v}}_2 = \bar{\mathbf{v}}_2^T \bar{\mathbf{v}}_3 = \bar{\mathbf{v}}_1^T \bar{\mathbf{v}}_3 = 0$.

3.3. Optimal Estimation of Vanishing Points

Assume that ${}^C\mathbf{m}_i, i \in \mathcal{M} = \{1, \dots, N\}$ are the moments of 3D lines in a Manhattan world that are observed by an intrinsically-calibrated camera. The noise-free normalized moments measured by the camera are denoted as ${}^C\bar{\mathbf{m}}_i$ and the noisy moment measurements as ${}^C\hat{\mathbf{m}}_i$. For now we assume that the lines are classified into three mutually orthogonal groups, and that no outlier exists among them. In Section 4, we describe how to relax these assumptions. Given a partitioning of \mathcal{M} into \mathcal{M}_j , $j = 1, 2, 3$ sets, each representing an orthogonal group, the optimal (in a weighted least-squares sense) values for the vanishing points on the Gaussian sphere are obtained by solving the following constrained least-squares problem:

$$\mathcal{P}_1 : \min_{\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \bar{\mathbf{v}}_3} \frac{1}{2} \sum_{j=1}^3 \sum_{i \in \mathcal{M}_j} \sigma_i^{-2} (\bar{\mathbf{v}}_j^T {}^C\hat{\mathbf{m}}_i)^2 \quad (1a)$$

$$\text{subject to } \bar{\mathbf{v}}_1^T \bar{\mathbf{v}}_2 = \bar{\mathbf{v}}_2^T \bar{\mathbf{v}}_3 = \bar{\mathbf{v}}_1^T \bar{\mathbf{v}}_3 = 0 \quad (1b)$$

$$\|\bar{\mathbf{v}}_1\| = \|\bar{\mathbf{v}}_2\| = \|\bar{\mathbf{v}}_3\| = 1 \quad (1c)$$

where σ_i are weights reflecting the uncertainty in each line-moment observation. The constraints in (1b)-(1c) can be satisfied, if we choose the vanishing points as the rows of a 3×3 orthonormal matrix. Thus, we can rewrite \mathcal{P}_1 as:

$$\mathcal{P}'_1 : \min_{\mathbf{C}} \frac{1}{2} \sum_{j=1}^3 \sum_{i \in \mathcal{M}_j} \sigma_i^{-2} (\bar{\mathbf{e}}_j^T \mathbf{C}^c \hat{\mathbf{m}}_i)^2 \quad (2a)$$

$$\text{subject to } \mathbf{C}^T \mathbf{C} = \mathbf{I}_3 \quad (2b)$$

where $\mathbf{C} = [\bar{\mathbf{v}}_1 \ \bar{\mathbf{v}}_2 \ \bar{\mathbf{v}}_3]^T$ is the unknown orthonormal matrix, and $\bar{\mathbf{e}}_1 = [1 \ 0 \ 0]^T$, $\bar{\mathbf{e}}_2 = [0 \ 1 \ 0]^T$, $\bar{\mathbf{e}}_3 = [0 \ 0 \ 1]^T$. Note that \mathbf{C} can be either proper with $\det(\mathbf{C}) = 1$ corresponding to a rotation matrix, or improper with $\det(\mathbf{C}) = -1$ corresponding to a reflection matrix. These two types of matrices can be converted to one another by negating one of their rows (or columns). We hereafter consider only the case of \mathbf{C} being a rotation matrix since $\bar{\mathbf{v}}_\ell$ and $-\bar{\mathbf{v}}_\ell$ correspond to the same vanishing points on the Gaussian sphere. In this case, \mathbf{C} is the camera's orientation in the Manhattan-world reference frame.

With the addition of $\det(\mathbf{C}) = 1$, \mathcal{P}'_1 [see (2)] becomes a special case of the orientation-from-line-correspondences problem. The general form of this problem is:

$$\mathcal{P}_2 : \arg \min_{\mathbf{C}} \frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} ({}^c \bar{\mathbf{n}}_i^T \mathbf{C}^c \hat{\mathbf{m}}_i)^2 \quad (3a)$$

$$\text{subject to } \mathbf{C}^T \mathbf{C} = \mathbf{I}_3, \det(\mathbf{C}) = 1. \quad (3b)$$

where ${}^c \bar{\mathbf{n}}_i$ is the *a priori* known 3D (not-necessarily-cardinal) direction of the i -th line in the global frame of reference and ${}^c \mathbf{C}$ is the rotation matrix representing the camera's orientation in the global frame of reference. This nonlinear weighted least-squares problem for $N \geq 3$ can be solved using iterative methods such as Gauss-Newton [11]. However, in the absence of an accurate initial estimate, iterative approaches may converge to local minima, and are not guaranteed to find all global minima. To address these limitations, we hereafter present an algebraic method that *directly* solves the nonlinear least-squares problem without requiring initialization.

We start by expressing the rotation matrix in (3) using the Cayley-Gibbs-Rodriguez (CGR) parametrization since (i) the components of the rotation matrix are naturally expressed as rational-polynomial functions of the CGR parameters, and (ii) CGR is a minimal representation of rotation, and thus, does not require additional constraints such as the ones in (3b) to ensure that it corresponds to a valid rotation [20]. Furthermore, the CGR parametrization introduces the minimum number of unknowns in the resulting polynomial system and hence allows fast computation of its solutions.

In CGR representation, a rotation matrix is expressed as

$$\mathbf{C}(\mathbf{s}) = \frac{\bar{\mathbf{C}}(\mathbf{s})}{1 + \mathbf{s}^T \mathbf{s}}, \quad \bar{\mathbf{C}}(\mathbf{s}) \triangleq ((1 - \mathbf{s}^T \mathbf{s})\mathbf{I}_3 + 2[\mathbf{s} \times] + 2\mathbf{s}\mathbf{s}^T).$$

where $\mathbf{s}^T = [s_1 \ s_2 \ s_3]$ is the vector of CGR parameters, and $[\mathbf{s} \times]$ is the corresponding skew-symmetric matrix. Substituting this in (3) yields:

$$\mathcal{P}'_2 : \arg \min_{\mathbf{s}} J, \quad J = \frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} ({}^c \bar{\mathbf{n}}_i^T \mathbf{C}(\mathbf{s})^c \hat{\mathbf{m}}_i)^2. \quad (4)$$

Compared to \mathcal{P}_2 [see (3)] the optimization constraint is now removed since the CGR parametrization ensures that $\mathbf{C}(\mathbf{s})$ is a rotation matrix. To algebraically find the global minimum of \mathcal{P}'_2 , we first determine all the critical points of J by solving the following optimality conditions, and then choose the one(s) that minimize \mathcal{P}'_2 . To derive the optimality conditions, we first factor out $(1 + \mathbf{s}^T \mathbf{s})$ from J in \mathcal{P}'_2 :

$$J = \frac{J'}{(1 + \mathbf{s}^T \mathbf{s})^2}, \quad J' \triangleq \frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} ({}^c \bar{\mathbf{n}}_i^T \bar{\mathbf{C}}(\mathbf{s})^c \hat{\mathbf{m}}_i)^2. \quad (5)$$

Then, the optimality conditions of \mathcal{P}'_2 are:

$$\frac{\partial J}{\partial s_j} = \frac{1}{(1 + \mathbf{s}^T \mathbf{s})^3} \left((1 + \mathbf{s}^T \mathbf{s}) \frac{\partial J'}{\partial s_j} - 4s_j J' \right) = 0 \quad (6a)$$

$$\frac{\partial J'}{\partial s_j} = \sum_{i=1}^N \sigma_i^{-2} ({}^c \bar{\mathbf{n}}_i^T \bar{\mathbf{C}}^c \hat{\mathbf{m}}_i) \frac{\partial}{\partial s_j} ({}^c \bar{\mathbf{n}}_i^T \bar{\mathbf{C}}^c \hat{\mathbf{m}}_i) \quad (6b)$$

for $j = 1, 2, 3$ and $N \geq 3$. Considering that $(1 + \mathbf{s}^T \mathbf{s})$ is nonzero for real \mathbf{s} , we simplify the optimality conditions as

$$f_j(\mathbf{s}) = (1 + \mathbf{s}^T \mathbf{s}) \frac{\partial J'}{\partial s_j} - 4s_j J' = 0, \quad j = 1, 2, 3. \quad (7)$$

These optimality conditions are fifth-order polynomials in the elements of \mathbf{s} whose real variety (i.e., solutions) comprises the critical points of \mathcal{P}'_2 . Directly solving these polynomials, however, is challenging since the ideal generated by them turns out to be non-zero dimensional, due to having a continuous variety on the imaginary hypersphere defined by $1 + \mathbf{s}^T \mathbf{s} = 0$.¹ To overcome this challenge, we introduce the following auxiliary polynomial that removes the (complex) solutions of $1 + \mathbf{s}^T \mathbf{s} = 0$ from the variety of $f_j = 0$:

$$f_0(s_0, \mathbf{s}) = s_0(1 + \mathbf{s}^T \mathbf{s}) - 1 = 0 \quad (8)$$

where s_0 is a new auxiliary variable. Note that this polynomial can be satisfied only if $1 + \mathbf{s}^T \mathbf{s}$ is nonzero. The

¹This is verified by computing the Hilbert dimension of the ideals generated by instances of the problem with integer or rational coefficients. Saturating these ideals with $1 + \mathbf{s}^T \mathbf{s}$ makes them zero dimensional.

new *saturated* system of polynomial equations consists of three fifth-order equations [see (7)] and one cubic equation [see (8)], in four unknowns (s_0, \mathbf{s}) . Assuming that at minimum three lines are observed, from which at least two are nonparallel, this polynomial system will have 40 solutions that can be computed by the method described in the following section. The globally optimal estimates are simply the solutions that minimize \mathcal{P}'_2 [see (4)].

Note that the computational complexity of solving the saturated polynomial system and finding the global minimum does not increase with the addition of measurements, since the degree and number of polynomials expressing the optimality conditions are fixed. Moreover, computing the contribution of all measurements to the coefficients of the polynomials f_j , $j = 1, 2, 3$ increases only linearly with the number of measurements.

3.4. Polynomial System Solver

Several methods exist for solving the polynomials describing the saturated optimality conditions of (7) and (8). Amongst them, numerical methods, such as Newton-Raphson, need initialization and may not find all the solutions. Symbolic reduction methods based on the computation of the system's Gröbner basis are capable of finding all roots without any initialization [8]. However, they can only be used for integer or rational coefficients since their application to floating-point numbers suffers from quick accumulation of round-off errors, which in turn, results in incorrect solutions [8]. Instead, we employ a method developed by Auzinger and Stetter [3] that computes the multiplication matrix (a generalization of the companion matrix to multivariate polynomial systems), whose eigenvalues are the roots of the associated polynomial system. In the following, we briefly describe an efficient method for computing the multiplication matrix.

Let us denote a *monomial* in $\mathbf{x} = [x_1 \cdots x_n]^T$ by $x^\gamma \triangleq x_1^{\gamma_1} x_2^{\gamma_2} \cdots x_n^{\gamma_n}$, $\gamma_i \in \mathbb{Z}_{\geq 0}$, with degree $\sum_{i=1}^n \gamma_i$. A polynomial of degree d in \mathbf{x} is denoted by $f = \mathbf{c}^T \mathbf{x}_d$ where \mathbf{x}_d is the $\binom{n+d}{n}$ -dimensional vector of monomials of degree up to and including d , and \mathbf{c} is the vector of coefficients of equal size. We assume that the given system of equations has n polynomials, denoted by $f_i = \mathbf{c}_i^T \mathbf{x}_{d_i} = 0$, $i = 1, \dots, n$, each of them with degree d_i . The total degree of the polynomial system is $d \triangleq \max_i d_i$. By padding the coefficient vectors of f_i 's with zeros, and stacking them together in \mathbf{C} , we can present the polynomial system in the matrix form of $\mathbf{C}\mathbf{x}_d = \mathbf{0}$.

A system of polynomial equations defines an *ideal* I as the set of all the polynomials that can be generated as $\sum_i f_i h_i$ where h_i is any polynomial in \mathbf{x} . Clearly the elements of the ideal become zero at the solutions of the original (generator) polynomial system. The Gröbner basis $G \triangleq \langle g_1, \dots, g_t \rangle$ of an ideal is a finite subset of the ideal

such that (i) the remainder of the division of any polynomial by it is unique, and (ii) any polynomial whose division by the Gröbner basis results in zero remainder, is a member of the associated ideal. The first property can be expressed as: $\varphi(\mathbf{x}) = r(\mathbf{x}) + \sum_{i=1}^t g_i(\mathbf{x})h_i(\mathbf{x})$ where φ is any polynomial in \mathbf{x} , h_i 's are the quotient polynomials, and r is the unique remainder. We hereafter use the name “remainder” as the remainder of the division of a polynomial by the Gröbner basis. The Gröbner basis for an ideal generated from polynomials with integer or rational numbers can be computed using implementations of the so-called Buchberger's algorithm [8] on symbolic software packages such as Macaulay2 or Maple. Computation of the Gröbner basis for polynomials with floating-point coefficients is much more difficult due to quick accumulation of round-off errors in the Buchberger's algorithm.

The remainders of the polynomials that are not in an ideal are instrumental in finding the solutions (i.e., variety) of that ideal. It can be shown that all such remainders can be expressed as a linear combination of a specific (unique) group of monomials that comprise the so-called *normal set* [8]. The normal set can be easily obtained from the Gröbner basis of an ideal, and under mild conditions,² its cardinality equals the number of solutions (real and complex) of the ideal, and it will contain the monomial 1 [8, p.43]. The important point here is that the normal set is generically fixed across different instantiations of the polynomials. Therefore, we can compute the normal set of an instance of the problem (e.g., using integer or rational coefficients) and use it when the coefficients are floating point.

Let us assume that the cardinality of the normal set is s , and represent its monomials in a vector form \mathbf{x}_B . Then multiplication of \mathbf{x}_B with a generic polynomial $\varphi(\mathbf{x})$ yields:

$$\varphi(\mathbf{x}) \cdot \mathbf{x}_B = \mathbf{M}_\varphi \mathbf{x}_B + \begin{bmatrix} h_{11} & \cdots & h_{1t} \\ \vdots & & \vdots \\ h_{s1} & \cdots & h_{st} \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_t \end{bmatrix} \quad (9)$$

where h_{ij} 's are polynomials in \mathbf{x} , and g_i 's are the elements of the Gröbner basis. In this expression, \mathbf{M}_φ is called the *multiplication matrix* associated with φ . This relationship holds since the remainder of any polynomial (including $x^\gamma \varphi(\mathbf{x})$, $x^\gamma \in \mathbf{x}_B$) can be written as a linear combination of elements in \mathbf{x}_B . Now, if we evaluate (9) at $\mathbf{x} = \mathbf{p}$, a solution of the ideal, all g_i 's become zero, and we obtain $\varphi(\mathbf{p}) \cdot \mathbf{p}_B = \mathbf{M}_\varphi \mathbf{p}_B$, where \mathbf{p}_B is \mathbf{x}_B evaluated at \mathbf{p} . Clearly, \mathbf{p}_B is an eigenvector of \mathbf{M}_φ , and $\varphi(\mathbf{p})$ is the associated eigenvalue. Therefore, if we select $\varphi(\mathbf{x})$ equal to one of the variables (e.g., x_i), we can read off the x_i -coordinate of the solutions as the eigenvalues of \mathbf{M}_φ . Furthermore, depending on the ordering of the monomials when computing

²These conditions are: (i) the ideal must be radical, (ii) its variety must be non-empty and zero dimensional. These conditions are generally satisfied for the vanishing-point estimation problem.

the Gröbner basis, \mathbf{x}_B may include all first-order monomials x_1, \dots, x_n . In that case, one can simultaneously read off all the coordinates of the solutions, for an arbitrary choice of φ , as long as it is nonzero and distinct at each solution of the ideal.

When the Gröbner basis can be computed (such as in polynomial systems with integer coefficients), one can use it directly to compute remainders of $\varphi(\mathbf{x}) \cdot \mathbf{x}_B$, and construct \mathbf{M}'_φ . This is not possible, however, when working with polynomials with floating-point coefficients. Therefore we employ the method proposed in [5] to compute \mathbf{M}'_φ . We first note that some of the monomials of $\varphi(\mathbf{x}) \cdot \mathbf{x}_B$ remain in \mathbf{x}_B , while some others do not. We form vector \mathbf{x}_R from the latter monomials, and write:

$$\varphi(\mathbf{x}) \cdot \mathbf{x}_B = \mathbf{M}'_\varphi \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_B \end{bmatrix} = \mathbf{M}'_{B,\varphi} \mathbf{x}_B + \mathbf{M}'_{R,\varphi} \mathbf{x}_R \quad (10)$$

where \mathbf{M}'_φ is called the unreduced multiplication matrix. We want to express the remainders of \mathbf{x}_R as a linear combination of \mathbf{x}_B without using the Gröbner basis. For this purpose, we expand each original polynomial f_i by multiplying it with all the monomials up to degree $\ell - d_i$ (ℓ to be determined later). Clearly all these new expanded polynomials belong to the ideal generated by the original polynomials, and they have monomials up to degree ℓ . Thus, we can write them collectively in matrix form as $\mathbf{C}_e \mathbf{x}_\ell = 0$. We reorder \mathbf{x}_ℓ and \mathbf{C}_e as:

$$\mathbf{C}_e \mathbf{x}_\ell = [\mathbf{C}_E \quad \mathbf{C}_R \quad \mathbf{C}_B] \begin{bmatrix} \mathbf{x}_E \\ \mathbf{x}_R \\ \mathbf{x}_B \end{bmatrix} = 0 \quad (11)$$

where \mathbf{x}_E are the monomials that belong neither to \mathbf{x}_R nor to \mathbf{x}_B . Multiplying (11) with \mathbf{N}^T , the left null space of \mathbf{C}_E , and decomposing $\mathbf{N}^T \mathbf{C}_R = \mathbf{Q}\mathbf{R} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$ using QR factorization, yields:

$$[\mathbf{N}^T \mathbf{C}_R \quad \mathbf{N}^T \mathbf{C}_B] \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_B \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \mathbf{R}_1 & \mathbf{Q}_1^T \mathbf{N}^T \mathbf{C}_B \\ \mathbf{0} & \mathbf{Q}_2^T \mathbf{N}^T \mathbf{C}_B \end{bmatrix} \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_B \end{bmatrix} = \mathbf{0}.$$

If ℓ is selected sufficiently large, \mathbf{R}_1 will be full rank [16], and we can conclude that $\mathbf{x}_R = -\mathbf{R}_1^{-1} \mathbf{Q}_1^T \mathbf{N}^T \mathbf{C}_B \mathbf{x}_B$. This is precisely the relationship that we need, in order to express \mathbf{x}_R in terms of \mathbf{x}_B . Substituting this relationship in (10) yields the multiplication matrix:

$$\mathbf{M}'_\varphi = \mathbf{M}'_\varphi \begin{bmatrix} \mathbf{I}_s \\ -\mathbf{R}_1^{-1} \mathbf{Q}_1^T \mathbf{N}^T \mathbf{C}_B \end{bmatrix}. \quad (12)$$

For solving equations (7) and (8), we had to expand the polynomials up to degree $\ell = 11$ and arrived at a multiplication matrix \mathbf{M}'_φ of dimensions 40×40 . Finally, we mention that it is possible to compute the multiplication matrix without explicit computation of the normal set. Further details on this subject and also on possible numerical instabilities and their remedies cannot be provided here due to space considerations. We refer the interested reader to [5, 16, 22].

3.5. Multiplicity of Solutions

To the best of our knowledge, no study is available that investigates the existence and determines the number of global minima of (4). However, there exist several studies of the conditions for having finite number of solutions to the *deterministic* rotation estimation problem, along with characterization of the maximum number of solutions. In noise-free situations, these conditions can be directly used to determine if (4) has a finite number of global minima (i.e., whether the *real* variety of the gradient ideal generated by (7) is zero dimensional), and in that case how many such global minima exist. Our simulations and experiments have confirmed that these results pertain in practical noisy scenarios.

The minimal requirement for estimating a camera's orientation from line observations is investigated by Chen [6]. In particular, he has shown that in order to have a finite number of solutions for the camera's orientation, three or more lines must be observed, and at least one of the 3D lines must be nonparallel with the others. In a Manhattan world, this corresponds to observing three lines that pass through at least two (out of three) vanishing points. In that case, Chen has shown that up to eight solutions for the camera's orientation may exist.

When more than three lines are observed, it is possible to reduce the number of valid solutions for the camera's orientation. However, following the method proposed in [15], it is easy to show that regardless of the number of measurements, there exist at least four solutions for a camera's orientation from observations of lines with known directions in a Manhattan world. Specifically, if we denote the 3D direction of the j th line in the global frame as $\mathbf{n}_j \in \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, $j = 1, \dots, N$ and their noise-free normalized moments measured in the camera's frame as ${}^c \mathbf{m}_j$, then, $\mathbf{n}_j^T {}^c \mathbf{C} {}^c \mathbf{m}_j = 0$, $j = 1, \dots, N$. Assuming that ${}^c \mathbf{C} = {}^c \mathbf{C}_1$ is one solution to this set of equations, it is easy to verify that ${}^c \mathbf{C} = \mathbf{\Pi} {}^c \mathbf{C}_1$ for

$$\mathbf{\Pi} \in \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

also satisfies the same set of equations. Clearly, any of the above choices for $\mathbf{\Pi}$ only reverses the direction of two of the vanishing points from $\bar{\mathbf{v}}_j$ to $-\bar{\mathbf{v}}_j$ (corresponding to two rows of ${}^c \mathbf{C}$). Therefore, we conclude that at most two distinct solutions for the orthogonal vanishing points can be obtained from observations of $N \geq 3$ lines passing through at least two vanishing points.

4. Classification of Lines

The *existing* approaches for RANSAC-based classification of lines use the intersection of two image lines (or their extensions) to generate hypotheses for *individual* vanishing

points [1, 23]. Since the vanishing points are detected sequentially (i.e., one after the other), at each step all lines are considered outliers unless they correspond to the dominant line direction. Once a dominant vanishing point is detected, all lines associated with it are removed from the image, and the procedure is repeated to detect the next dominant vanishing point. Besides assuming the existence of a dominant direction, these methods require a large number of hypotheses to compensate for the lines that do not pass through the dominant vanishing point (even if they are along cardinal directions). Moreover, the vanishing points determined in this way are not generally orthogonal.

In this work, we propose a more efficient and robust RANSAC-based approach that generates hypotheses for all three orthogonal vanishing points at once. Specifically, we randomly sample triplets of lines, and consider four possible configurations for their directions: one configuration assumes that each line is along one cardinal direction, and three configurations assume that two lines (out of three) are along one cardinal direction, while the third line is along another cardinal direction. Then, given these four possible configurations for each triplet, we employ the method described in Section 3.3 to compute at most eight hypotheses for all three orthogonal vanishing points (at most two hypotheses per configuration - see Section 3.5). After processing several sample triplets (typically 10), we will have M hypotheses for the three orthogonal vanishing points, denoted as $\bar{\mathbf{v}}_{i,\ell}$, $i = 1, 2, 3, \ell = 1, \dots, M$. We measure the angle between the j th line-moment, ${}^c\hat{\mathbf{m}}_j$, from each vanishing point as $\sin^{-1}(\bar{\mathbf{v}}_{i,\ell}^T {}^c\hat{\mathbf{m}}_j)$. If this angle is smaller than a prespecified threshold for any of the three orthogonal vanishing points of the ℓ th hypothesis, we label the j th line as inlier with respect to the ℓ th hypothesis. In this case, within the ℓ th hypothesis, we classify the j th line as belonging to the vanishing point that generated the smallest angle. The winner of the RANSAC algorithm is the hypothesis that results in the largest number of inliers.

Note that triplets that do not produce any valid hypotheses are the ones that include line(s) with non-Manhattan directions, or only parallel lines. Clearly, the set of such triplets is significantly smaller than the set of samples generating invalid hypotheses for the intersection-based RANSAC classifiers [1, 23]. Therefore the presented method requires considerably fewer samples to generate at least one valid hypothesis. Additionally, this method does not assume any particular dominant direction, and directly provides all the orthogonal vanishing points.

When the computational resources are limited, the algorithm described in Section 3.3 may be too slow to generate the RANSAC hypotheses. In these situations, one may employ the deterministic minimal solver proposed in [6] to generate hypotheses. In the presence of noise, however, this method may return complex-valued solutions whose real

parts are far from the minimizers of \mathcal{P}'_2 [see (4)] (this is in contrast to the method described in Section 3.3 that always finds the real-valued minimizers of \mathcal{P}'_2). In such cases, one would need to take more triplet samples to compensate for the potential loss of valid hypotheses whose deterministic solutions are complex-valued. Alternatively, we may minimize a *relaxed* optimization problem (described in the Appendix) that only requires solving three cubic polynomials (at a fraction of time of the non-relaxed problem), and always returns real-valued solutions. Our simulations (not provided here due to space limitations) have shown that the minimizers of this relaxed problem accurately approximate those of the original non-relaxed cost function.

5. Experiments

We have tested the proposed algorithms on the outdoor and indoor images from the York Urban Database (YUD) [9]. We have compared three sets of vanishing-point estimates: **GT**: The ground truth provided in the database; **ALS**: The analytical least-squares estimates (Section 3.3) using the labeled ground-truth lines in the database; **RNS**: The analytical least-squares estimates (Section 3.3) using lines extracted from the images, and classified using the proposed RANSAC algorithm (Section 4). For both the **ALS** and the **RNS** methods, we have used the intrinsic camera calibration parameters provided by the YUD.

For the **RNS** method, we have employed Canny's edge detector followed by Kovese's edge linking and line-segment fitting [13] to extract straight-line segments from each image (see Fig. 1). The number of detected segments was chosen equal to the number of ground-truth lines in the database to ensure fair comparison of the methods. The number of sample triplets for the RANSAC algorithm was adaptively selected and it was typically between 4 to 15 samples [11]. In order to determine the candidate orthogonal vanishing points from each triplet, we used the algorithm described in the Appendix, whose C++ implementation takes about 4 ms to find the solutions on a Core 2 Duo processor. and the minimum angle for labeling a moment measurement as inlier was set to 2.5° .

In both **ALS** and **RNS**, we employed the algorithm described in Section 3.3 to obtain the optimal estimates for the orthogonal vanishing points from the labeled line segments. For this purpose, we expanded the polynomials up to $\ell = 13$. The resulting \mathbf{C}_e is a sparse 2860×3060 matrix with less than 3% fill-ins. Our current C++ implementation of the symbolic-numeric method described in Section 3.4 takes ~ 290 ms to find the solutions.

The results of the **RNS** for each image of the YUD along with the C++ code is available at <http://umn.edu/~faraz/vp>. Here, in Fig. 1, we show a few of the **RNS** results, to demonstrate its performance in the presence of outliers. Among the 102 images provided in the database,

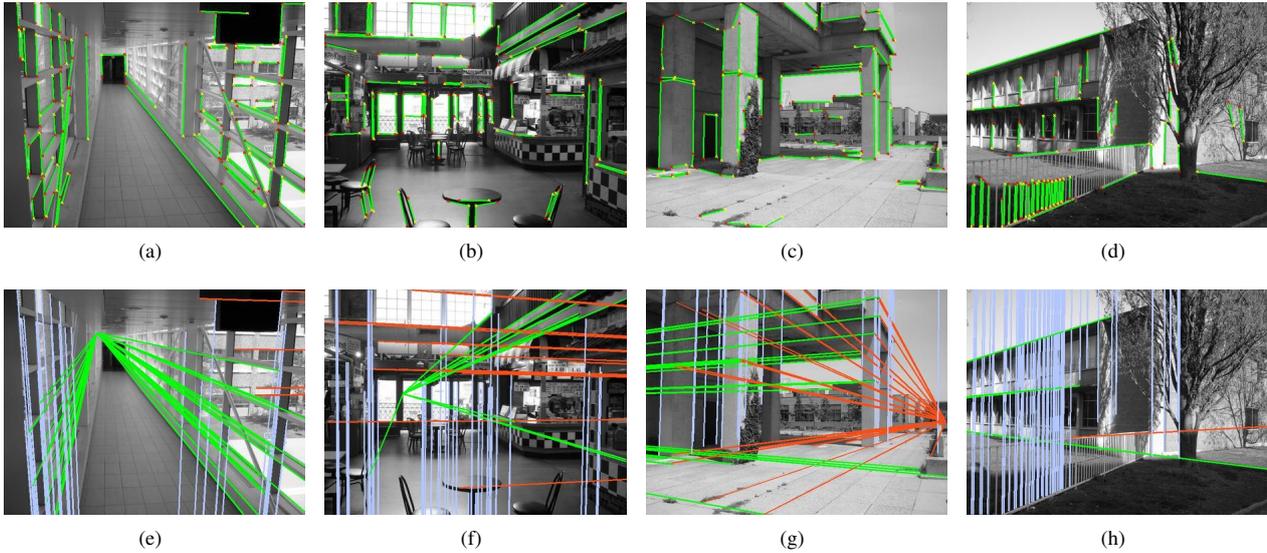


Figure 1. A selection of the results from the York Urban Database (YUD) [9]. (a,b,c,d): The extracted line segments using Canny's edge detector, and Kovese's line segmentation algorithm [13]; (e,f,g,h): The estimated orthogonal vanishing points using the **RNS** method. The complete set of the results is available in the supplied additional material.

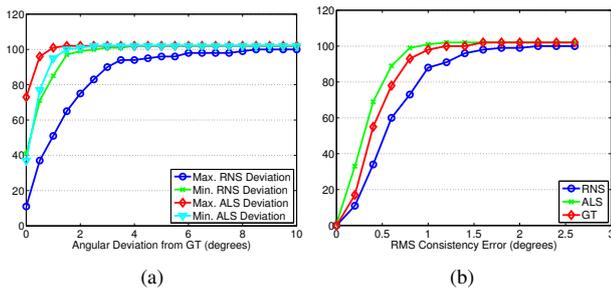


Figure 2. Cumulative histogram of (a): The angular deviation from the Ground Truth, and (b): RMS consistency error.

the **RNS** failed only once, due to the existence of many non-Manhattan directions. The maximum and minimum angular deviation of the vanishing points estimated by the **ALS** and the **RNS**, compared to the ground truth, **GT**, is provided in Fig. 2(a). Using the **RNS**, only two images result in an angular deviation larger than 10° . Note that although the number of line segments given to the **ALS** and **RNS** is the same, the latter may include an unknown number of outliers, while the former only contains labeled ground-truth lines. Thus, the slightly inferior results of the **RNS** are to be expected since it uses fewer (inlier) lines to estimate the orthogonal vanishing points.

Finally, we demonstrate the consistency of the estimated vanishing points achieved by each method in Fig 2(b). The consistency error of each (inlier) line segment is defined as $\sin^{-1}(\hat{\mathbf{v}}_i^T \mathbf{c} \hat{\mathbf{m}}_j)$, where $\hat{\mathbf{v}}_i$ is the corresponding estimated vanishing point. The RMS consistency error is computed

across all inlier line segments of an image. Using the **RNS**, 100 of the images result in RMS consistency errors less than 3° . Interestingly, the RMS consistency error of the **ALS** is even lower than the **GT**. This can be explained by the fact that the **GT** vanishing points were first estimated individually from hand-labeled lines (i.e., without imposing orthogonality constraints), and then an orthogonal set of vanishing points were fit to them [9]. This possibly leads to suboptimal estimation of orthogonal vanishing points by the **GT**.

6. Conclusion

In this work, we present a new method for analytically estimating the optimal (in the least-squares sense) orthogonal vanishing points in a Manhattan world. Specifically, we employ the multiplication matrix to solve the multivariate polynomial system resulting from the optimality conditions of the corresponding constrained least-squares problem and compute all its critical points. Amongst these, the ones that minimize the cost function are the globally optimal estimates of the orthogonal vanishing points. Additionally, we introduce a robust and efficient RANSAC-based line classifier that employs the optimal estimator to generate hypotheses for all three orthogonal points from triplets of line observations.

We are currently extending this method to simultaneously estimate the camera's focal length along with the orthogonal vanishing points. Estimating several orthogonal triplets of vanishing points in an "Atlanta World" [18] is part of our future work.

A. Relaxed Estimation of Vanishing Points

We may relax the original problem \mathcal{P}_2 in (3) by requiring \mathbf{C} to be only orthogonal and not necessarily orthonormal. The relaxed problem is

$$\mathcal{P}_3 : \underset{\mathbf{C}, \alpha}{\arg \min} \frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} (\mathbf{c}^T \hat{\mathbf{n}}_i^T \mathbf{C}^c \hat{\mathbf{m}}_i)^2 \quad (13a)$$

$$\text{subject to } \mathbf{C}^T \mathbf{C} = \alpha^2 \mathbf{I}_3, \det(\mathbf{C}) \geq 1. \quad (13b)$$

This is equivalent to relaxing (1c) as $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}_3\| \geq 1$. Although \mathbf{C} is not a rotation matrix, a valid rotation matrix can be easily obtained as $\bar{\mathbf{C}}/\alpha$. Parametrizing $\bar{\mathbf{C}}$ using CGR parameters yields:

$$\mathcal{P}'_3 : \underset{\mathbf{s}}{\arg \min} J' \quad (14)$$

where J' is defined in (5), and $\alpha = 1 + \mathbf{s}^T \mathbf{s}$. The cost function J' is a 4th-order polynomial in the elements of \mathbf{s} . To algebraically find the global minimum of (14), we first determine all the critical points of J' by solving the optimality conditions $\frac{\partial J'}{\partial s_j} = 0$ for $j = 1, 2, 3$ and $N \geq 3$ [see (6b)]. These three relaxed optimality conditions are always *cubic* polynomials, regardless of the number of measurements, and generally generate a zero-dimensional ideal with 27 solutions that can be obtained from the corresponding multiplication matrix. Among these solutions (critical points of J'), we choose the ones that minimize J as the relaxed estimates for the globally optimal solutions of \mathcal{P}'_2 .

References

- [1] D. G. Aguilera, J. G. Lahoz, and J. F. Codes. A new method for vanishing points detection in 3D reconstruction from a single view. In *Proc. ISPRS Commission V Workshop Virtual Reconst. Visual. Complex Archit.*, Mestre-Venice, Italy, Aug. 22–24 2005. **1, 2, 6**
- [2] M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 282–289, Hilton Head, SC, June 13–15 2000. **1**
- [3] W. Auzinger and H. J. Stetter. An elimination algorithm for the computation of all zeros of a system of multivariate polynomial equations. In *Proc. Int. Conf. on Numer. Math.*, pages 11–30, Singapore, 1988. **4**
- [4] S. T. Barnard. Interpreting perspective images. *Artif. Intell.*, 21(4):435–462, Nov. 1983. **1**
- [5] M. Byröd, K. Josephson, and K. Åström. A column-pivoting based strategy for monomial ordering in numerical Gröbner basis calculations. In *Proc. Euro. Conf. Comput. Vision*, pages 130–143, Marseille, France, Oct. 12–18 2008. **5**
- [6] H. H. Chen. Pose determination from line-to-plane correspondences: existence condition and closed-form solutions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(6):530–541, June 1991. **5, 6**
- [7] J. M. Coughlan and A. L. Yuille. Manhattan world: Orientation and outlier detection by bayesian inference. *Neural Comput.*, 15(5):1063–1088, May 2003. **1**
- [8] D. Cox, J. Little, and D. O’Shea. *Using Algebraic Geometry*. Springer, 2004. **4**
- [9] P. Denis, J. Elder, and F. Estrada. Efficient edge-based methods for estimating Manhattan frames in urban imagery. In *Proc. Euro. Conf. Comput. Vision*, pages 197–210, Marseille, France, Oct. 12–18 2008. **1, 6, 7**
- [10] W. Förstner. Optimal vanishing point detection and rotation estimation of single images from a legoland scene. In *Proc. ISPRS Commission III Symp. Photogramm. Comput. Vis. Image Anal.*, pages 157–162, Sept. 2010. **2**
- [11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. **3, 6**
- [12] J. Košecká and W. Zhang. Video compass. In *Proc. Euro. Conf. Comput. Vision*, pages 29–32, Copenhagen, Denmark, May 27 - June 2 2002. **1**
- [13] P. D. Kovesi. MATLAB and Octave functions for computer vision and image processing. University of Western Australia. Available from: <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>. **6, 7**
- [14] E. Lutton, H. Maitre, and J. Lopez-Krahe. Contribution to the determination of vanishing points using Hough transform. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(4):430–438, Apr. 1994. **1**
- [15] N. Navab and O. Faugeras. Monocular pose determination from lines: critical sets and maximum number of solutions. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 254–260, New York City, NY, June 1993. **5**
- [16] G. Reid and L. Zhi. Solving polynomial systems via symbolic-numeric reduction to geometric involutive form. *J. Symb. Comput.*, 44(3):280–291, Mar. 2009. **5**
- [17] C. Rother. A new approach to vanishing point detection in architectural environments. *Image Vis. Comput.*, 20(9–10):647–655, 2002. **1**
- [18] G. Schindler and F. Dellaert. Atlanta world: an expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 203–209, Washington, DC, June 27 – July 2 2004. **1, 7**
- [19] J. Shufelt. Performance evaluation and analysis of vanishing point detection techniques. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(3):282–288, Mar. 1999. **1**
- [20] M. D. Shuster. A survey of attitude representations. *J. Astronaut. Sci.*, 41(4):439–517, Oct.–Dec. 1993. **3**
- [21] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *Proc. Int. Conf. Comput. Vision*, pages 1250–1257, Kyoto, Japan, Sept. 29 – Oct. 2 2009. **1, 2**
- [22] N. Trawny, X. S. Zhou, and S. I. Roumeliotis. 3D relative pose estimation from six distances. In *Proc. Robot. Sci. Syst.*, Seattle, WA, June 28– July 1, 2009. **5**
- [23] H. Wildenauer and M. Vincze. Vanishing point detection in complex man-made worlds. In *Proc. Int. Conf. Image Anal. Process.*, pages 615–622, Modena, Italy, Sept. 2007. **1, 2, 6**